Can test list context manipulations improve recognition accuracy in the DRM paradigm?

Raymond W. Gunter, Stacey L. Ivanko, and Glen E. Bodner

University of Calgary, Alberta, Canada

Only test-based manipulations can be used to help people distinguish accurate from false memories once events have been encoded. In two experiments we examined how the type of studied words (weak vs strong associates, or less vs more memorable associates) and nonstudied lure words (related vs unrelated lures) on the test list affect recognition accuracy in the Deese-Roediger-McDermott paradigm. False recognition of critical lures decreased substantially in the related-lure context, but so did correct recognition of studied words. False recognition was little affected by the studied-word manipulations. In general, participants claimed to recognise critical lures as often as weak associates or less memorable studied words but less often than either strong associates or more memorable studied words. The test-list context affected how participants classified their recognition experiences but it did not systematically change their overall memory accuracy.

Although memory is often accurate, distortions and illusions of memory can and do occur (see Roediger, 1996, for a review). People report having experienced events that never occurred (Hyman, Husband, & Billings, 1995), recall or recognise details of events that were implanted after the events occurred (Loftus, Miller, & Burns, 1978; Loftus & Palmer, 1974), and claim to remember words (Underwood, 1965) or pictures (Koutstaal & Schacter, 1997) that were never presented. It is important to distinguish false from true memories when evaluating eyewitness testimony, delivering psychotherapeutic services, and conducting forensic interviews (e.g., Lindsay & Read, 1994), and many researchers have attempted to identify conditions under which memory errors can be reduced (see Bruce & Winograd, 1998, for a historical review). The present experiments examine how one variable—the recognition test-list context—influences how people classify their recognition experiences in the Deese-Roediger-McDermott (DRM) paradigm.

The DRM paradigm is often used to create memory illusions in the lab (Deese, 1959; Roediger & McDermott, 1995). Participants study lists of thematically related words (e.g., note, sound, piano, sing). Each list converges on a single associate that is not presented on the list, dubbed the critical lure (e.g., music). Deese (1959) found that participants falsely recalled the critical lures far more often than other nonstudied words. Roediger and McDermott (1995) further showed that participants recalled the critical lures about as often as studied words from the middle of the study lists. They also extended the illusion to the domain of recognition memory, and found that participants claimed to recognise the critical lures as often as the particular selection of studied words that were included on the recognition test (i.e., from list positions 1, 8, and 10).

Roediger and McDermott (1995) also asked participants to make remember/know judgements about each recognised word (Gardiner, 1988; Rajaram, 1993; Tulving, 1985).
involves the recollection of episodic details about the presentation of an item at study (e.g., a distinctive memory of a thought, image, or reaction). Knowing refers to recognition experiences in which an item seems familiar in the absence of any specific recollections. About half of the critical lures received remember judgements, indicating that participants often experience specific illusory recollections for the critical lures rather than merely vague feelings of familiarity.

ACCOUNTS OF THE DRM ILLUSION

Several accounts of the DRM illusion have been proposed (e.g., Brainerd, Reyna, & Poole, 2000; McDermott & Watson, 2001; Roediger & McDermott, 1995). In an early account, Underwood (1965) argued that critical lures could become activated during encoding via implicit associative responses (see also Roediger & McDermott, 1995). Critical lures may come to mind consciously (e.g., Goodwin, Meissner, & Ericsson, 2001) or their representations in a semantic network may be activated unconsciously (Seamon, Lee, Toner, Wheeler, Goodkind, & Birch, 2002a) via spreading activation (e.g., Collins & Loftus, 1975). At test, participants can misattribute either experience to the critical lures having been presented, although Seamon et al. (2002a) reported that conscious generation of critical lures at study does not predict false recognition.

More recently, McDermott and Watson (2001) proposed a dual-process activation-monitoring account in which an activation-based process was supplemented by a strategic control process termed monitoring. In this account, use of the monitoring process at study and/or test (i.e., keeping track of which words were actually presented) should reduce false recognition. Another dual-process account, fuzzy trace theory, postulates that two types of memory trace are deposited in parallel during study and that either/both can be used to guide recognition judgements at test (e.g., Brainerd & Reyna, 1996, 1998; Brainerd et al., 2000; Brainerd, Wright, Reyna, & Payne, 2002). Verbatim traces contain the identities of individual list items. Gist traces contain the overall meaning or theme of the items on a particular list. Use of verbatim-based traces should enable the rememberer to realise that critical lures were not studied (e.g., Brainerd et al., 2000; Payne, Elie, Blackwell, & Neuschatz, 1996). In contrast, reliance on gist-based traces can produce “phantom recollection” of the gist-consistent critical lures (e.g., Brainerd, Wright, Reyna, & Mojardin, 2001).

INFLUENCES OF ENCODING AND TEST MANIPULATIONS ON THE DRM ILLUSION

Many studies have explored the effect of encoding manipulations on the DRM illusion (see Roediger, McDermott, & Robinson, 1998, for a review). Some of these manipulations attenuate false memory, including multiple list presentations (McDermott, 1996; Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002b), increasing the distinctiveness of the list words (Dodson & Schacter, 2001; Israel & Schacter, 1997), and blocking list words in various configurations (Goodwin et al., 2001). Other encoding manipulations increase false memory, including dividing attention (Perez-Mata, Read, & Diges, 2002) and semantic encoding of list words (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001).

Fewer studies have examined the influence of test-based manipulations on false memory. This is surprising, given that memory performance is a function of both encoding and retrieval factors (Tulving, 1972), and given that only test conditions can be controlled once events have been encoded. It is therefore especially useful on a practical level to look for test-based manipulations that can help individuals distinguish accurate from false memories, despite admonitions that the DRM illusion may be beyond conscious control (Gallo, Roediger, & McDermott, 2001; Seamon et al., 2002a). Whittlesea (2002) found that false recognition increased when memory for critical lures (e.g., doctor) was tested in semantically incongruent (vs congruent) sentence contexts (e.g., “She was miserable from the damp and the … doctor”). Other studies have instructed participants to forget each studied list after its presentation, prior to a recall test (the list method of directed forgetting). These instructions reduced recall of studied words, but had little effect on false recall (Seamon, Luo, Shulman, Toner, & Cagler, 2002) or even increased it (Kimball & Bjork, 2002). Recently, Marsh, McDermott, and Roediger (2004) examined the role played by test-based priming in creating false recall and recognition. They found that the number of related words tested before a critical lure had little impact on the DRM illusion.
Perhaps the most common approach to eliminating false recognition at test has been to warn one group of participants about the DRM illusion after the study phase and to compare their recognition to a non-warned group. Gallo, Roberts, and Seamon’s (1997) “cautious” group were warned that some words on the recognition test would be similar to words heard at study but were not actually presented, and they were urged to avoid false recognition. This warning reduced correct recognition but did not reduce false recognition. Warning also failed to decrease false recognition in Anastasi, Rhodes, and Burns (2000). Finally, Gallo, McDermott, Percer, and Roediger (2001) did find a decrease in false recognition after a warning, but the warning reduced correct recognition as well.

At best, warnings given before the recognition test have had limited success at reducing the DRM illusion. Moreover, giving participants explicit warnings places pressure on them to respond conservatively. A drawback to such an approach is that this pressure may encourage participants to mistrust or to outright betray their internal recognition experiences. Participants may feel compelled to respond “no” to critical lures (and studied words) for which they subjectively experience some recognition. Instructional pressures may increase the likelihood that participants will respond “no” to the critical lures without actually altering the illusion of memory they subjectively experience.

TEST-LIST CONTEXT AND FALSE RECOGNITION

The goal of our experiments was to examine whether a test-based manipulation other than explicit instructions might improve participants’ ability to distinguish studied words from critical lures, without externally pressuring them to betray their internal recognition experiences. We wanted to see whether situations that encourage careful evaluation of one’s recognition experiences might reduce the DRM illusion relative to situations that do not. The approach we adopted was simply to manipulate the nature of the studied and lure words presented on the recognition test list.

The composition of the test list has recently been shown to have a substantial effect on recognition of studied words. Bodner and Lindsay (2003) found that participants made more remember judgements to words studied in a medium-level-of-processing task when these items were mixed at test with words studied in a shallow rather than in a deep level-of-processing task. When asked to report the basis for their remember judgements, the likelihood that participants reported particular bases for remembering the medium items depended on the test-list context. The medium-with-deep group were more likely to base their remember judgements on the list in which an item was studied, whereas the medium-with-shallow group were more likely to base their remember judgements on thoughts and associations they experienced during study. Bodner and Lindsay concluded that the test-list context can influence participants’ functional definitions of remembering and knowing (see also Gruppuso, Lindsay, & Kelley, 1997).

Our experiments investigate whether two aspects of the test-list context can also modulate false recognition. First, we manipulated whether the lure words (i.e., nonstudied items other than the critical lures) were related or unrelated to the list themes. It was expected that related lures would be recognised at a higher rate than unrelated lures, of course, but the key question was how this manipulation would influence false recognition of the critical lures. Given that related lures should be more difficult to discriminate from both studied words and critical lures, we reasoned that the presence of related lures should discourage participants from simply endorsing any words consistent with the gist of a studied list. Moreover, some research indicates that recognition experiences for critical lures involve less perceptual information and fewer contextual associations (e.g., reactions triggered by the item at study) than recognition experiences for studied words (Mather, Henkel, & Johnson, 1997; Norman & Schacter, 1997). Thus, the DRM illusion should be reduced if the related-lure context allows participants to discover and use diagnostic differences in their recollective experiences for studied and nonstudied words to guide their recognition decisions.

Our second test-list manipulation departed from the tradition of evaluating correct recognition using studied words from positions 1, 8, and 10 of each studied list, begun by Roediger and McDermott (1995). Instead, in Experiment 1 the three strongest (positions 1 to 3) or the three weakest associates (positions 10 to 12) from each
12-item study list were tested. Analogous to the logic of the lure manipulation, we reasoned that it might be more difficult to distinguish critical lures from the strong associates than from the weak associates. For example, critical lure words may co-occur more often with strong associates than with weak associates, making them more confusable. If so, a test list containing studied words that are all strong associates of the critical lures might also encourage participants to respond more carefully, and to seek out aspects of their recollective experiences that will maximise discrimination between studied and nonstudied items.

**EXPERIMENT 1**

Using a factorial manipulation, Experiment 1 examined how the DRM illusion is influenced by the type of lures (related vs unrelated) and the type of studied words (strong vs weak associates) on the recognition test. Remember/know judgements were collected to reveal how the test-list context influences subjective aspects of participants’ recognition experiences.

**Method**

**Participants.** Undergraduate students from the University of Calgary participated for extra course credit \((n = 83)\). Three of these participants were excluded because they accurately reported the purpose of the experiment (see procedure), leaving 19–21 participants per group.

**Materials.** A total of 20 lists from Roediger and McDermott (1995) were used. Words from positions 3, 8, and 13 were trimmed from each list to serve as the lures on the recognition test. Study lists were divided into two sets of 10 such that each set produced similarly potent DRM illusions based on the norms of Stadler, Roediger, and McDermott (1999). Each 12-item list was recorded on audiotape (approximately one word every 1.5 seconds) with a short pause between lists.

The 70-word recognition test consisted of a single sheet of paper with an underlined space beside each word. Ten of the words were the critical lures. The 30 studied words were either the strongest associates (from list positions 1–3) or the weakest associates (from list positions 10–12) of the critical lures. The 30 lure words were either related (the 3 trimmed items from each studied list) or unrelated (the 3 trimmed items from each nonstudied list) to a list theme. Studied word type and lure type were factorially varied across four groups. Because two sets of study lists were used, two recognition sheets were required for each group. The 70 items were arranged in a random order and the type of studied word or lure in a given list position was swapped across the four conditions. The critical lures were presented in the same 10 list positions across the four conditions.

**Procedure.** Participants were tested in sessions containing up to 10 people. They were told that they would hear several lists of words and that each list would be separated by a pause. A tape recorder was used to play back one of the two sets of 10 lists. Each study list set was heard by approximately half of the participants in each group of the design. Recognition sheets were then handed out such that a similar number of participants served in each test-list context within each session. Participants were instructed to decide whether each word on the test had been presented on one of the study lists or not. If the word was recognised, participants were instructed to classify their recognition experience as an experience of remembering or of knowing by marking an R or K in the space beside the word. Participants were asked to leave the space beside words that were not recognised blank. The distinction between remember and know was explained verbally and appeared on top of the recognition sheet for reference. After the recognition test, participants were given a questionnaire to determine if they knew or had figured out the purpose of the experiment.

**Results**

The mean proportion of remember and know (hereafter R and K) judgements for lures, studied words, and critical lures in Experiment 1 are presented in Figure 1. The sum of R and K judgements represents overall recognition. List set did not affect the pattern of results and was not included in the analyses reported below. The three dependent measures (overall recognition, R judgements, and K judgements) were analysed separately. We also performed analyses based on hit-minus-false-alarm scores and A’ scores but these analyses are not reported because they did not
Figure 1. Mean proportion of remember and know judgements for lures, studied words, and critical lures in Experiment 1 as a function of whether the recognition test list contained weak associate (W) or strong associate (S) studied words and unrelated (U) or related (R) lure words.

...affect our conclusions regarding the effectiveness of the test-list manipulations; these analyses are more important when separate baselines for nonstudied critical lure and nonstudied list words are used. Judgements of knowing were not much influenced by list context, hence for brevity we report only the effects of our manipulations on overall recognition and R judgements to lures, studied words, and critical lures. The significance level was set at .05.

Recognition of lures and studied words. Recognition of lures and studied words was analysed using separate one-way analyses of variance (ANOVA)s because different items were tested depending on the level of each factor. The manipulation of lure type was successful; false alarms were more likely for related lures than for unrelated lures overall (.18 vs .12), \( F(1, 78) = 4.68, MSE = .02 \), and for R judgements (.06 vs .03), \( F(1, 78) = 5.50, MSE = .004 \). As for the manipulation of studied words, strong associates were more likely to be recognised than weak associates overall (.71 vs .61), \( F(1, 78) = 7.67, MSE = .03 \), but they were not more likely to receive R judgements (.42 vs .39), \( F < 1 \).

Recognition of critical lures. Participants falsely recognised more than 60% of the critical lures on average, confirming that our experiment produced a robust DRM illusion. More importantly, Figure 1 suggests that the size of this illusion depended on the test-list context. False recognition of the critical lures was analysed using 2 (studied word type: strong vs weak) × 2 (lure type: related vs unrelated) between-groups ANOVA, because the same critical lures were present in each cell of the design. The two factors did not interact in either dependent measure (\( F < 1 \)), so only their main effects are considered below.

The presence of strong versus weak studied associates had no effect on overall recognition of the critical lures (.64 vs .65), \( F < 1 \). The trend towards fewer R judgements in the strong-
associate context relative to the weak-associate context did not reach significance (.31 vs .40), $F(1, 76) = 2.69, MSE = .06, p = .11$. In contrast, the lure type manipulation had a pronounced effect on the DRM illusion. Critical lures were less likely to be recognised in the related-lure context than in the unrelated-lure context overall (.57 vs .72), $F(1, 76) = 7.12, MSE = .06$, and were also less likely to receive R judgements (.27 vs .44), $F(1, 76) = 8.90, MSE = .06$. The related-lure context seems to have induced participants to evaluate their recognition experiences more carefully, resulting in a sizeable decrease in false recognition of the critical lures.

Recognition of critical lures vs studied words. Examination of Figure 1 suggests that the related-lure context did not selectively reduce false recognition of critical lures—it also reduced correct recognition of studied words. One-way ANOVAs revealed that correct recognition was lower in the related-lure context than in the unrelated-lure context (.62 vs .70), $F(1, 78) = 5.64, MSE = .03$, as was the rate of R judgements (.35 vs .46), $F(1, 78) = 7.68, MSE = .03$. Participants thus appear to have become more cautious but not more accurate in the presence of related lures.

To further assess this possibility, recognition of studied words and critical lures was compared using 2 (studied type: strong vs. weak) $\times$ 2 (lure type: related vs unrelated) $\times$ 2 (item type: studied words vs critical lures) mixed-factor ANOVAs. The three-way interactions were not significant, $Fs < 1$. Item type did not interact with lure type overall, $F(1, 76) = 1.74, MSE = .02, p = .19$, or for R judgements, $F(1, 76) = 1.50, MSE = .02, p = .22$, confirming that recognition of studied words and critical lures decreased to a similar extent in the related-lure context. Item type did interact with studied type overall, $F(1, 76) = 5.68, MSE = .02$, and for R judgements, $F(1, 76) = 6.78, MSE = .02$. Follow-up one-way ANOVAs revealed that critical lures were less likely to be recognised than strong associates (.64 vs .71), $F(1, 38) = 4.13, MSE = .03$, and were less likely to receive R judgements (.31 vs .42), $F(1, 37) = 9.90, MSE = .03$. In contrast, critical lures were as likely to be recognised as weak associates overall (.65 vs .61), $F(1, 40) = 1.64, MSE = .02, p = .21$, and were as likely to be given R judgements (.40 vs .39), $F < 1$. Participants could differentiate between critical lures and strong associates, but not between critical lures and weak associates.

Discussion

At first glance, the decrease in the DRM illusion in the related-lure context relative to the unrelated-lure context appears to be the most striking result in Experiment 1. The effect of presenting related lures on the test list was comparable in size to the effect of providing explicit warnings after study (e.g., Gallo et al., 2001), and has the advantage of not externally pressuring participants to respond “no” even when they are potentially experiencing some familiarity and/or recollection for an item. However, turning to correct recognition, a pitfall of test-based attempts at reducing false recognition becomes obvious, namely that the drop in correct recognition in the related-lure context was on par with the drop in false recognition. Drops in correct recognition also occurred in some warning studies (e.g., Gallo et al., 1997, 2001).

Although strong associates were more likely to be recognised than weak associates, the studied word context did not affect the DRM illusion, perhaps because the rates of R judgements for weak and strong associates (and hence the richness of participants’ set of recollective experiences) were the same in both contexts. In Experiment 2 we used a different manipulation of the studied word context; our goal was to give participants a richer set of remembering experiences for the studied words in one context than in another. To accomplish this goal, we simply selected more versus less memorable studied words using the R judgement data from Experiment 1. The more-memorable studied word context should provide a rich set of recollective experiences that might allow participants to realise that “something is missing” in their recognition experiences for the critical lures. In fuzzy-trace theory terms, the presence of more memorable words should help cue verbatim traces, which should reduce false recognition (e.g., Brainerd, Reyna, & Kneer, 1995).

**EXPERIMENT 2**

Experiment 2 was a replication of Experiment 1 using a different manipulation of the studied words included on the recognition test list. Here we examined whether the presence of less versus more memorable studied words would influence false recognition of the critical lures.
Method

Participants. Another 84 students from the Experiment 1 pool were randomly assigned to one of the four test-list contexts (n = 21 each).

Materials and procedure. The rates of remembering for the weak and strong associates in Experiment 1 were used to categorise the studied words on each list as less versus more memorable. We then selected the 10 lists with the largest difference in remember judgements between the less-memorable and the more-memorable words. The mean rate of R judgements was .21 for the less-memorable words (range .05 to .45) and was .64 for the more-memorable words (range .38 to .85). This set of lists, recorded as in Experiment 1, was presented to all participants. The recognition test lists were also constructed as in Experiment 1, except each weak or strong associate was replaced with a less- or more-memorable studied word. Thus, the 30 studied words were either the 3 most memorable or the 3 least memorable words of the 6 words from each list for which recognition data were collected in Experiment 1. The 30 lures were either the 3 related words trimmed from each of the 10 studied lists (related lures), or the 3 related words trimmed from the 10 nonstudied lists (unrelated lures). Because the same 10 lists were presented to all participants, only four different recognition sheets (one for each test-list context) were needed. Otherwise, the materials and procedure were as in Experiment 1.

Results

The mean proportion of R and K judgements for lures, studied words, and critical lures are presented in Figure 2 and were analysed as in Experiment 1.

Recognition of lures and studied words. As expected, the more-memorable studied words were better recognised than the less-memorable studied words overall (.73 vs .61), F(1, 82) = 13.50, MSE = .02, and were also more likely to be remembered (.52 vs .38), F(1, 82) = 11.69, MSE = .03. Fortuitously, the lure type manipulation for the 10 lists used in Experiment 2 was also more effective than in Experiment 1. Related lures were much more likely to be recognised than unrelated lures overall (.25 vs .07), F(1, 82) = 36.03, MSE = .02, and were substantially more likely to receive R judgements (.11 vs .02), F(1, 82) = 37.40, MSE = .01.

Recognition of critical lures. The lure type manipulation again had a sizeable influence on the DRM illusion. Critical lures were less likely to be recognised in the related-lure context than in the unrelated-lure context (.56 vs .71), F(1, 80) = 9.08, MSE = .05, and were less likely to receive R judgements (.33 vs .46), F(1, 80) = 6.37, MSE = .05. However, as in Experiment 1, recognition of the critical lures was not influenced by the studied word context provided on the recognition test. Overall false recognition of the critical lures was not reduced in the more-memorable context relative to the less-memorable context (.61 vs .66), F < 1.01. There were somewhat fewer R judgements to critical lures in the more-memorable context than in the less-memorable context (.36 vs .43), but this drop did not approach significance, F(1, 80) = 1.71, MSE = .05, p = .19. The interaction of lure type and studied type was not significant for either dependent measure, Fs < 1.

Recognition of critical lures vs studied words. As in Experiment 1, recognition of both studied words and critical lures decreased in the related-lure context (see Figure 2). Analyses confirmed that correct recognition decreased in the related-lure context relative to the unrelated-lure context (.63 vs .71), F(1, 82) = 5.38, MSE = .02, as did the rate of R judgements (.38 vs .52), F(1, 82) = 12.36, MSE = .03.

Next, we compared the effects of the two manipulations on recognition of studied words versus critical lures, as in Experiment 1. The drop in recognition in the related-lure context was similar for studied words and critical lures, as reflected in a non-significant interaction between item type and lure type on overall recognition, F(1, 80) = 2.14, MSE = .03, p = .15, and on R judgements, F < 1. Also replicating Experiment 1, item type interacted with studied type in the overall recognition measure, F(1, 80) = 11.96, MSE = .03, and for R judgements, F(1, 80) = 21.49, MSE = .02. The three-way interactions were not significant, Fs < 1. Follow-up ANOVAs revealed that critical lures were less likely to be recognised than more-memorable studied words overall (.61 vs .73), F(1, 41) = 16.80, MSE = .02, and in terms of R judgements (.36 vs .52), F(1, 41) = 34.66, MSE = .01. In contrast, participants were equally likely to claim to recognise critical lures and less-
memorable studied words overall (.66 vs .61), \( F(1, 41) = 1.46, MSE = .03, p = .23 \), and in terms of R judgments (.43 vs .38), \( F(1, 41) = 2.25, MSE = .03, p = .14 \). Analogous to the pattern found in Experiment 1, participants could differentiate between critical lures and more-memorable studied words, but not between critical lures and less-memorable studied words.

**Discussion**

In Experiment 2, the recognition test-list context consisted of related or unrelated lures and less or more memorable studied words, and the results largely replicated Experiment 1. First, even though the memorability of the studied words was successfully manipulated in Experiment 2, the studied word context did not affect the DRM illusion. Second, the DRM illusion was substantially reduced in the related-lure context, but correct recognition was equivalently reduced. The related-lure context made participants interpret their recognition experiences more cautiously, but it did not improve their ability to discriminate between studied words and critical lures.

**GENERAL DISCUSSION**

The recognition test list context in most studies of the DRM illusion consists of studied words from positions 1, 8, and 10 of each list along with a set of unrelated lures. We departed from this tradition by examining the effect of different types of studied words and lure words on the DRM illusion. We presented either the strongest or weakest associates of the critical lure from each studied list (Experiment 1) or the least or most memorable words from each studied list (Experiment 2; based on the data from Experiment 1) on the test list. Recognition of critical lures was not much affected by either of these manipulations of the studied words. The inclusion of related versus unrelated lures on the test list, on the other hand, had a robust effect in both experiments: Critical lures
were much less likely to be recognised and remembered in the related-lure context. If the test-list context did not matter, the DRM illusion should have been similar in all test-list conditions, given that all participants studied the same lists, were tested on the same critical lures, and received the same recognition instructions. The lure-type results fit well with Bodner and Lindsay's (2003) finding that the test-list context can affect how participants classify their recognition experiences.

Although false recognition decreased in the related-lure context, so too did correct recognition of studied words. The inclusion of related lures may have confused participants ("They all seem familiar. I can’t tell which words were presented and which were not.") or otherwise made them reluctant to endorse any word related to a list theme. The effect of related lures was similar to the effect of warning participants about the DRM illusion just prior to the test (e.g., Anastasi et al., 2000; Gallo et al., 1997, 2001): Participants were more cautious but were not better able to distinguish critical lures from studied words. According to the activation-monitoring account, cueing use of the monitoring process at test by requiring participants to make finer discriminations between studied and nonstudied words should have reduced false recognition. However, the ineffectiveness of test-based warnings led McDermott and Watson (2001) to suggest that monitoring at study may typically be more effective than monitoring at test. The present results concur with this suggestion. In the remainder of this discussion, we further evaluate our test-list manipulations, and close by suggesting other approaches to improving recognition memory accuracy in this paradigm.

**Evaluating test-based manipulations**

**Studied-word manipulations.** A notable new finding here is that participants were less likely to recognize critical lures than either strong associates (Experiment 1) or more-memorable words (Experiment 2), but were as likely to recognize critical lures as either weak associates or less-memorable words. The choice of studied words included on the recognition test therefore affects the conclusion drawn regarding whether participants can distinguish critical lures from studied words. Our results highlight the arbitrariness of using studied words from list positions 1, 8, and 10 as a benchmark for assessing the relative memorability of critical lures versus studied words (e.g., Roediger & McDermott, 1995).

The relative memorability of the studied words in Experiment 2 did not have a reliable effect on the DRM illusion. Although Brainerd et al. (1995) found a reduction in false recognition by cueing verbatim memories at test through the presentation of memorable studied items, at least one study has failed to replicate this effect (Wallace, Malone, Swiergosz, & Amberg, 2000). Our findings stand with those of Wallace et al. (2000) and do not support the notion that reinstating verbatim traces at test better enables participants to edit out false recollections (see also Marsh et al., 2004). However, it is also possible that our manipulation of memorability was too weak; R judgements were only .14 more likely for more-memorable than for less-memorable studied words. Although neither studied-word manipulation reliably influenced the DRM illusion, pooling the two experiments revealed a small but significant decrease in R judgements to the critical lures in the strong-associate/more-memorable context (.34 vs .42), $F(1, 162) = 4.07$, $MSE = .06$. Thus, the nature of the studied words on the recognition test list likely exerts some influence on false recognition. Refinements of our studied-word manipulations might prove more successful.

**Lure-word manipulations.** We hypothesised that the related-lure context would lead participants to evaluate their recognition experiences more carefully, thereby enabling them to discover and use differences in their recollective experiences for studied and nonstudied words to guide their recognition responses. An increased reliance on recollection should reduce false recognition and bolster correct recognition. However, the related-lure context may instead have decreased reliance on gist without increasing reliance on recollection. Decreased reliance on gist should reduce both correct and false recognition (given that both studied words and critical lures are related to the list themes). The observed decrease in both correct and false recognition is consistent with a decreased use of gist that was not offset by an increased use of recollection.

Alternatively, the similarity of the drops in correct and false recognition in the related-lure context in Experiments 1 and 2 make it tempting to suggest that participants’ recollective experi-
ences for critical lures and studied words are inseparable, or at least that test-based manipulations cannot lead participants to detect differences in their recognition experiences for these two item types. Given that recollective experiences for studied words and critical lures do differ on some phenomenological characteristics (e.g., Mather et al., 1997; Norman & Schacter, 1997), a potentially effective way of reducing the DRM illusion at test might be to give participants a set of diagnostic criteria for distinguishing between true and false recognition experiences.

To date, only one study has examined whether the provision of potentially diagnostic criteria enables participants to avoid false recognition. Neuschatz, Payne, Lampinen, and Toglia (2001) read participants a list of phenomenological differences (e.g., perceptual, emotional, contextual) between true and false memories after the study phase and instructed them to use these differences to avoid false alarms during the test. These manipulations did not reduce false recognition, although participants’ confidence was somewhat reduced. However, the list of criteria was rather long and participants were not able to refer to it during the recognition test, so it is unclear how well participants retained or attempted to use this information. It is possible that information about diagnostic phenomenological differences might be more effective in conjunction with a related-lure context that encourages careful responding.

**Encoding-retrieval manipulations**

Our motivation for seeing whether test-based manipulations alone could reduce the DRM illusion was that in most applied situations, encoding has already transpired, making test-based manipulations the only tools available for trying to increase people’s memory accuracy. Unfortunately, the evidence to date suggests that test-based manipulations have limited success at selectively reducing false recognition (Experiments 1 and 2; Anastasi et al., 2000; Gallo et al., 1997, 2001; Marsh et al., 2004). A more effective strategy for improving memory accuracy will likely require consideration of both encoding and test conditions. Specifically, increasing item-specific processing of list words at study, combined with cueing reliance on this distinctive processing at test, may improve correct recognition while attenuating the DRM illusion (e.g., Gallo et al., 2001; Maylor & Mo, 1999).

Item-specific and relational processing accounts of recognition (Arndt & Reder, 2003; Einstein & Hunt, 1980) posit that encoding of relational (i.e., gist) information enhances both veridical and false recognition, while encoding of item-specific information (e.g., perceptual details associated with individual words, or verbatim information) enhances only veridical recognition (see also Schacter, Israel, & Racine, 1999). Furthermore, a number of studies have found that providing participants with distinctive item-specific information for each word at study (e.g., a picture) causes participants to adopt a specific decision-making rule at test, a distinctiveness heuristic. Participants using a distinctive heuristic are able to avoid false recognition by treating the presence/absence of specific aspects of their recollective experiences for individual words (e.g., memory for the picture presented at study) as diagnostic regarding an item’s old/new status. Although our test-based manipulations were intended to lead participants to evaluate their general recollective experiences for individual words more carefully, we did not provide participants with specific features for each word at study that could be used to formulate a helpful decision rule. In contrast, providing distinctive item-specific information at study has been found to selectively reduce false recognition while not affecting correct recognition (see Dodson & Schacter, 2002).

The benefits of encoding contexts that foster item-specific processing can also be combined with test contexts that encourage participants to use this information to guide their recognition judgements. Reinstating distinctive item-specific information at test should encourage participants to make use of a specific rule like the distinctiveness heuristic. Several studies have found that reinstating item-specific study cues at test decreases false recognition (e.g., Israel & Schacter, 1997; Schacter et al., 1999). Moreover, under some conditions a match between study and test conditions has been shown to benefit correct memory performance (e.g., Dean, Yerkovich, & Gray, 1988). It should therefore be possible to use knowledge about the original encoding conditions to develop memory test contexts that will facilitate people’s ability to weed out false memories, while preserving or even enhancing their ability to recollect what actually happened.
REFERENCES


